A Short Note on Prior Work Relevant to Concatenated Context-Selected Linear Mappings (CCSLM)

CCSLM — routing among a bank of linear maps per layer via a separate context channel, with the main data channel staying purely linear along any single realized path — sits at the intersection of a few older lines of work. None of them is exactly this, but each anticipates a piece of it.

Switched linear systems (control theory). The oldest relevant idea: a system that behaves linearly at each instant, but which linear dynamics apply is chosen by a switching signal, often exogenous or state-triggered. Stability isn't governed by any single matrix's eigenvalues but by the *joint spectral radius* of the whole matrix set — some switching sequences can destabilize a system even when every individual mode is stable. This is the right lens for asking whether an adversarial or noisy context sequence could blow up your network's norm across depth.

Gated Linear Networks (Veness et al., 2017–2021, DeepMind). The closest architectural cousin. Each neuron holds several weight vectors and a *hard, data-dependent gate* — typically a half-space partition of a side-information vector — selects which weight vector is used, with no nonlinearity elsewhere. Backprop is discarded in favor of local, per-neuron online convex optimization, which is what buys the paper's strong universality result. The key structural parallel to CCSLM: context and data are explicitly separate channels, and the nonlinearity of the whole system emerges entirely from *which* linear map fires, not from any activation function.

Context-dependent gating / XdG (Masse et al., 2018). From continual learning, not from a "make gating expressive" motivation. A random, fixed binary mask is generated per task and used to zero out ~80% of hidden units, so different tasks touch mostly disjoint parts of the network. Relevant here as the degenerate case where context is a static task label rather than something computed from input, and where gating is unlearned scaffolding rather than a trained routing function.

Mixture-of-Experts / hypernetworks. MoE learns a *soft or top-k* router over a bank of expert subnetworks; hypernetworks generate an entire weight matrix from a context vector rather than selecting among a fixed bank. CCSLM sits between them — a discrete/binary selector like MoE's routing, but over a small fixed bank of purely linear experts, closer in spirit to a hypernetwork with a highly restricted (lookup-table) output space rather than a generative one.

Piecewise-linear / spline theory of deep nets (Montufar et al.; Balestriero & Baraniuk). ReLU networks are, formally, piecewise-linear functions where the input itself determines the linear region via activation patterns. CCSLM generalizes this by *decoupling* the partition-selector from the thing being transformed — context need not be a function of x at all (weather, a task ID) — trading the geometry of hyperplane arrangements for whatever geometry the context function induces (e.g., LSH bucket boundaries).

The throughline: every one of these lines picks a different point on the same two axes — (1) is the router learned or fixed, and (2) is the router's input the same as the data being transformed, or a separate channel. CCSLM is a fairly clean, unclaimed point in that space: learned-or-fixed context, always channel-separated, with hard binary selection composing into literal matrix products across depth.